3 NETHERLANDS

The Evolution of Achievement Gaps from Early Childhood to Adolescence in the Netherlands

Thomas van Huizen
Utrecht University

3.1 Introduction

School segregation is comparatively high and rising in the Netherlands (Ladd et al. 2011; Boterman 2019; Inspectorate of Education 2018; Education Council 2018). An explanation for the high degree of social and ethnic segregation is the strong commitment to "freedom of education", safeguarded in the Dutch constitution for over a century. The education system is characterized by a high degree of parental choice, equal funding of public and private schools and a high degree of school autonomy. In the first part of the previous century there were no major concerns about social and ethnic segregation as religious denomination rather than social class or ethnicity determined school choice. However, increasing secularisation and the influx of migrants in the 1960's and 1970's resulted in higher levels of school segregation. Instead of reducing segregation per se (e.g. by limiting parental choice), the chosen policy objective was to provide high quality education for all children by alleviating the negative effects of segregation (Ritzen et al. 1997; Ladd & Fiske 2011). Since 1985 more public resources are allocated to schools with a higher share of disadvantaged children according to a weighted funded system. Moreover, substantial investments have been made in targeted preschool programs to alleviate skill disadvantages before school entry (Akgündüz & Heijnen 2018; Leseman et al. 2017). Hence, the Dutch education system aims to neutralize the key arguments against segregation by allocating additional resources to schools with high concentrations of disadvantaged children and by providing disadvantaged families access to relatively high-quality preschool programs.

The Dutch 'high segregation – high compensation' system is often considered successful in terms of efficiency and equity of achievement results: "The Dutch school system is one of the best in the OECD, as measured by the Programme of International Student Assessment (PISA) and the Survey of Adult Skills (PIAAC). It is also equitable, with a very low proportion of poor performers." (OECD 2016b: p.11). This result is especially striking because spending on education is not exceptionally high, indicating that the Dutch education system is rather efficient.[1] Recent international evidence based on PIRLS shows that in the Netherlands a comparatively small share of the variation in Grade 4 reading scores can be explained by school differences (UNICEF 2018). This tentatively suggests that the high degree of school segregation is effectively counteracted by the weighted funded system. Moreover, while SES gaps in Grade 4 reading scores in the Netherlands are higher than those in most Scandinavian countries, they are lower than many continental European countries (e.g. Germany, France, Belgium) (Rözer & van de Werfhorst 2017: p.55). However, several policy concerns remain: segregation may have negative

[1] Spending primary education is low (measures as a percentage of GDP) to average (measures as spending per student). Spending on secondary education is above the OECD average though (OECD 2018b).

consequences beyond achievement scores, e.g. in terms of social. Moreover, students are tracked in secondary school from around the age of 12. Although the jury is still out on the effects of tracking, limited permeability between educational tracks is an important policy issue (Inspectorate of Education 2018; Education Council 2018).

This chapter examines the evolution of achievement inequalities from early childhood to adolescence in the Netherlands. The central aim of this study is to complement the cross-sectional snapshots of achievement gaps by providing an answer to both research questions discussed in Chapter 1 of this report. First, I test when SES and migration-related gaps in achievement arise and how these gaps evolve during (early) childhood and adolescence. Are gaps already sizeable before children enter kindergarten and school? Do gaps narrow or widen as children move through primary and secondary school? Second, I examine the extent to which social and migration-related achievement gaps in primary school can be attributed to preschool achievement differences. Is there a substantial additional role of SES and migration background in the school years or are inequalities already settled in the years before school entry? Are low SES children with the same initial achievement as their high SES peers being left behind in primary school? Is there significant upward or downward mobility in achievement during the school years and does this vary by SES and migration background?

The results presented in this Chapter are based on two longitudinal datasets: Pre-COOL and COOL. Both datasets contain information on family SES and migration background. SES is measured by the level of education of the parents. Unfortunately, limited information about household income available. Moreover, migration inequalities are examined by estimating migrant-native gaps. In addition to overall migrant-native gaps, I examine whether specific migrant groups lag more behind than others, focusing on children with a Turkish or Moroccan background (two ISOTIS target groups).

COOL and Pre-COOL include data on a battery of achievement tests. The analysis focuses on inequalities in language/literacy and math/numeracy achievement as these two domains have been tested consistently in (Pre-)COOL. Moreover, from a life course (human capital) perspective these skills are highly relevant as language and math skills measured in childhood/adolescence significantly predict adult earnings and employment prospects (Chetty et al. 2014; Lin et al. 2016). Importantly, by combining the two related datasets, the evolution of gaps can be examined over an extended observation window (from age 2 to age 14).

Notwithstanding the strengths of the data, some serious limitations remain. First, as in Chapter 2, the analysis of the evolution of gaps relies on multiple cohorts. The data does only allow genuine longitudinal analysis from age 2 to 6 and from age 5 to 11 (and from 8 to 14). In line with the analysis presented in Chapter 2, a weighing strategy is employed to link different cohorts. Moreover, I discuss results that indicate that cohort effects in this time period are probably negligible. A second limitation is that, whereas COOL is a representative sample, Pre-COOL is not designed to be representative. Yet, by exploiting the overlap between the two datasets (i.e. kindergarten achievement results), weights can be assigned to correct for non-representativeness of the Pre-COOL sample. When comparing COOL and Pre-COOL results, the estimates appear to be very consistent with each other. Finally, the major limitation is probably panel attrition. Although longitudinal weights are used to limit concerns associated with non-random attrition, the sample size for longitudinal analysis is relatively small as a consequence of panel attrition. This implies that the study of research question 2, explaining school gaps by preschool differences, is limited to SES and overall migrant inequalities and does not distinguish between different migration groups.

The remaining of the chapter is structured as follows. The next section provides an overview of the Dutch institutional setting. In the third section, the data, measures and methodology to address the main research questions are discussed. Subsequently, the fourth section presents the empirical results. The final section concludes and discusses several policy implications.

3.2 Institutional context

3.2.1 Family policies

With respect to family policies, the Netherlands stands out from other continental European countries. For instance, in most continental European countries public spending for families with children is substantially above the OECD average, whereas it is below average in the Netherlands (Thévenon 2011). Maternity and parental leave policies are not generous for European standards: post-natal maternity leave is three months (100% replacement rate) and the duration of parental leave is 6 months. However, parental leave entitlement is on a part-time basis and it is often unpaid (the level of payment, if any, is regulated in the collective labour market agreements). Furthermore, full-time childcare attendance is rare in the Netherlands (see 3.2.2). Hence, the Netherlands shares some features with the UK, combining short leave with part-time childcare services.

A unique feature is that The Netherlands is “the first part-time economy in the world” (Visser 2002: p.23) and this is reflected in the rather unique solution to reconcile work and family life (Plantenga 2002). Female (maternal) employment is relatively high, but around three out of four working women – not just mothers with young children – are employed on a part-time basis. Given the short leave duration, it is common for mothers in the Netherlands to return to employment within the first half year after childbirth, but only for a limited number of hours. In the years before the child enters kindergarten at age four, the typical solution is part-time employment of mothers (reduction of working hours of fathers is also not uncommon), around two days of formal day care and one day of informal care by grandparents.

3.2.2 Early Childhood Education and Care

3.2.2.1 Participation in ECEC

Given the ‘Dutch part-time solution’, most Dutch children spend some time in ECEC before entering kindergarten. ECEC participation rates in centre-based care and regulated family day care are high for children below age three: 56% versus the EU average of 32%. Only Denmark has a higher ECEC participation rate for this age group (OECD 2016a).[2] As in all other EU countries, participation rates are higher for children aged three to five (92% in the Netherlands). Because kindergartens are free and universally available (see Section 3.2.2), participation jumps from below 81% for children aged three to over 96 for children aged 4 (at age 5 participation is

[2] As in most EU countries, ECEC participation rises sharply with age: whereas 23% of children aged 0 participates in (centre/family) day care, this increases to 56% for children aged 3 (Statistics Netherlands 2016).

mandatory).

In the Dutch case the ECEC attendance rates do not provide an accurate indication of the total exposure (intensity x duration) to ECEC as children are typically enrolled in childcare for a limited number of days per week. The modal use of centre day care is two days a week and the average ECEC hours is around 16 hours for the 0–2 group (which is comparable to the UK). In fact, only a small percentage of Dutch children spend 30 hours or more in ECEC: 6% of children aged 0–2 and 14% of children aged 3–4 (OECD 2016a). This implies that in terms of full-time equivalents, the ECEC participation rate of 0/2-year-olds is below the EU average (32.8 versus 36.7). Hence, ECEC participation rates for children below age three are among the highest in EU but below the EU average in the full-time equivalents.

As in many European countries, the Dutch ECEC system is organized as a split system, with different arrangements for children until age four and for children aged four until the start of primary school. For children below the age of four, various child care arrangements exist; some services are aimed at supporting work and family life and encouraging female labour market participation (day care centres), whereas others have typically a stronger focus on education (playgroups). An important feature of the Dutch educational system is that it has specific policies to reduce educational disadvantages at an early age.

3.2.2.2 The Dutch ECEC system: Alternatives, eligibility and funding

Day care centres (*kinderdagverblijven*) provide childcare services for zero- to four-year-olds and are the most commonly used facility for children in this age range. Centres provide full-day care for five days a week and primarily offer services to dual-earner families and are for that reason under the responsibility of the Ministry of Social Affairs and Employment. Since the introduction of the 2005 Child Care Act (*Wet Kinderopvang*), the day care market has been privatized and both commercial (for profit) and non-profit organizations operate on the market (see Noailly & Visser (2009) and Akgündüz & Plantenga (2014a) for a more extensive discussion of the reform). The Dutch day care system stands out as one of the few European countries without public provision of day care. The financing system is demand-driven: parents can select their preferred centre, conditional on availability of slots, and receive an income-dependent subsidy through the tax system. Although prices are not strictly capped, there is a soft price cap as there is a maximum hourly fee that is subsidized (7.18 euros in 2017). Day care prices are therefore generally not substantially higher than this maximum price. The subsidy system has been reformed several times since 2005; initially subsidies increased, resulting in drop in average parental costs from 37% in 2005 to 18% in 2007 and, consequently, an expansion of the day care sector: public spending on childcare subsidies tripled over the period 2004–2009 to 3 billion euro (Bettendorf et al. 2015). The subsidies were reduced somewhat in 2009 and more substantially in 2012. During the more recent years, subsidies have become more generous again. In the recent period, parents paid around one third of the gross price: net prices for the lowest income groups are 6%, whereas higher income groups pay up to two thirds of the gross price.[3] Subsidies for the second child in day care are substantially higher. It should be noted that actual day care costs for parents are rather uncertain ex ante, as the subsidies are based on realized (ex post) annual household income. Importantly, families are eligible for child care subsidies when both parents in the

[3] These figures refer to 2017. Given that subsidies are set for a maximum hourly price, actual parental costs may be higher.

household are employed, in education or actively looking for work; single-parent families are also eligible when the parent is employed, in education or actively looking for work. However, breadwinner families are not eligible for day care subsidies.

As an alternative to day care centres, working parents can also family day care services (*gastouderopvang*). Although less popular than day care centres, family day care services are a nontrivial part of the Dutch ECEC landscape, with around 7% of children aged 0–3 participating in these arrangements. Family day care is also regulated and subsidized and can be considered as formal, non-centre based ECEC. In case these facilities are registered in the national child care register, the same subsidy conditions apply as those for day care centres. Fees for in-home care services are generally lower than fees for day care centres.

In addition to day care centres and in-home services that provide full-day care for zero-to four-year-olds, playgroups (*peuterspeelzalen*) offer a part-day and more formal type of ECEC for children between age two and half and age four (before entering kindergarten). For children in the relevant age range, playgroups represent an important alternative to centre/family day care: 28% participates in playgroups, compared to the 43% in day care (and 9% in both types) (Statistics Netherlands 2015). Playgroups are run by municipal welfare organizations, day care centres and primary schools. Playgroups only provide half-day programs (2.5 hours per day), typically two days per week for around 40 weeks per year. Given the limited operating hours, the aim of playgroups is not to facilitate parental employment but rather to prepare children for kindergarten and school. Instead of subsidies through the tax system, playgroups are subsidized by municipalities (supply subsidies) and parents pay an income-dependent fee determined by the municipality.[4] Since the implementation of the Day Care and Playgroups Harmonization Act in January 2018, there are no formal differences between day care and playgroup organizations and the latter as a type of organization no longer exist. However, the type of service (i.e. an ECEC program of two half-days) are still offered by day care organizations, including the organizations formerly registered as playgroup. During the years before the 2018 reform there was a significant decline in the number of preschools.[5] In fact, most preschools formally changed from preschools to day care centres, providing de facto the same service but allowing dual-earner to receive subsidies for preschool services.

Furthermore, next to day care centres, in-home care services and playgroups that are in principle universally accessible, an important feature of the Dutch ECEC system is that it includes preschools targeted towards disadvantaged children. Since the 1970's, the Netherlands introduced policies that aim to actively reduce educational disadvantages. In the current educational setting, the Dutch Educational Disadvantage Policy (*Onderwijsachterstandenbeleid*) includes pre- and early school programs (*Voor- en Vroegschoolse Educatie*; *VVE*): preschool programs for children aged two-and-a-half to four, and early school programs for four- and five-year-olds. The latter are part of kindergarten and are discussed below. The central goal of these targeted preschools is to reduce early educational and developmental disparities. The preschool programs consist of 10–12 hours (four half-days) centre-based ECEC per week for around 40 weeks.[6] Programs are provided by day care centres or (before the 2018 reform) playgroup

---

[4] For instance, in 2018 in Utrecht the annual parental fees varied between 45 and 790 euro, depending on the level of household income.

[5] The number of preschool organizations dropped by almost 40% between the end of 2013 and early 2017 (Buitenhek, 2017).

[6] Municipalities are legally responsible for the provision of preschool programs of (at least) 10 hours per week for

organizations. The national budget is allocated between municipalities using the primary school weights system. Essentially, municipalities receive a larger part of the budget when they have more primary schools with a high proportion of children with low educated parents. The total preschool budget has increased considerably during the last decade, from 200 million euros to 260 million euros in 2011. Furthermore, substantial additional funding was allocated to the 37 largest municipalities (70 million in 2012 and 95 million in 2013) (Akgündüz & Heijnen 2018). Municipalities use this funding to provide subsidies to centres offering preschool programs and have a certain degree of autonomy concerning their preschool policies. In most municipalities eligible children can participate in preschool for free or for a small parental fee. Municipalities also differ in their targeting (eligibility) criteria, although the educational background of parents generally plays a major role.

3.2.2.3 Inequality issues

The actual use of ECEC is strongly related to family SES. There is a significant difference in the use of ECEC services for 0–2 aged children between families with higher and lower educated mothers: 46% (no tertiary education) versus 67% (with tertiary education) (OECD 2016a). Concerning 2/3-year-olds, available data indicates that ECEC participation increases substantially by household income: 92% of children in the highest household income quintile participate in ECEC versus 62% in the lowest income quintile, i.e. almost 4 out of 10 children in the lowest income group do not participate in ECEC before entering kindergarten (Statistics Netherlands 2015).

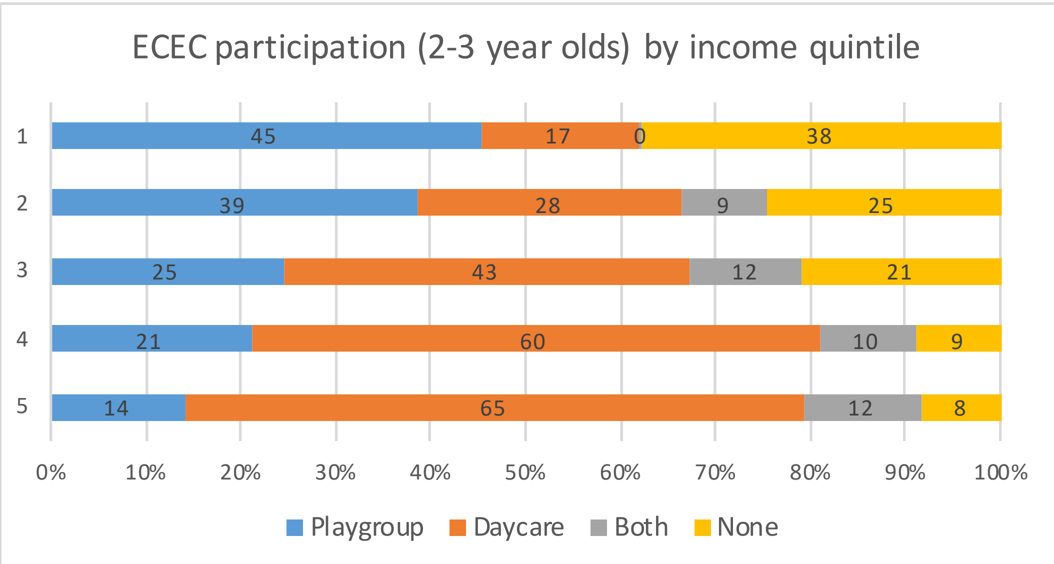

Figure 1 ECEC participation (2-3 years olds) by income quintile.

Source: Statistics Netherlands (2015).

disadvantaged children.

Families with higher income are not only more likely to use ECEC services, they also opt for different types of ECEC services. As shown in Figure 1, the propensity to use playgroups declines with household income, whereas the propensity to use day care increases with household income. This can be expected as the lowest income households are often not eligible for subsidies; these families are generally breadwinner families or families with no employed parent).

Empirical findings based on Pre-COOL show that equality of ECEC quality is to a large extent achieved in the Netherlands. First, preschools offer higher quality than day care centres. Second, ECEC providers with a higher share of disadvantaged children provide higher structural and process quality: "These results indicate that the targeted preschool policy in the Netherlands succeeds in providing higher quality ECEC to those who need it most" (Leseman et al. 2017: p.182). Results from multivariate analyses that take into account selection into ECEC generally find no major gaps in ECEC quality accessible to different groups (Akgündüz & Plantenga 2014b). Overall, these results are somewhat more mixed. On the one hand, higher SES families seem to use ECEC services with higher emotional support. On the other hand, native families use services of lower emotional support compared to migrant families.

While there is no causal evidence available on the impact of participation of targeted preschools, some results tentatively suggest that targeted preschools effectively reduce disadvantages. First, results based on Pre-COOL indicate that disadvantaged children catch up significantly compared to non-disadvantaged children in vocabulary and selective attention (executive function). However, it is unclear whether this can be attributed to participation in preschools. As Leseman and colleagues (2017: p.186) note: "The design of the pre-COOL study does not allow for strong conclusions about the effectiveness of participating in centre-based ECEC for developmental and educational outcomes. The catching-up effects that were found cannot be attributed unambiguously to participation in ECEC because no meaningful comparison could be made with equivalent children without any ECEC experience." Second, Akgündüz & Heijnen (2018) follow a quasi-experimental approach (difference-in-differences), exploiting a reform that increased funding for targeted preschools substantially for the 37 largest Dutch municipalities. The results show that grade repetition in kindergarten – an indicator for school readiness – declined by .8–1.8 percentage points for the targeted population. This effect is substantial given a grade retention rate of around 10% for the specific group.

3.2.3 Kindergarten and primary school

3.2.3.1 Central features

Dutch primary schools provide education from 4- to 12-year old children and consist of 8 'groups' (years), where group 1 and 2 provide a kindergarten program and are often mixed in terms of age. Formal schooling in primary schools starts in group 3 (around age 6; Grade 1). While in many other European countries kindergartens operate to a large extent independently from primary schools, in the Netherlands kindergartens are completely integrated in the primary school system. This implies that in the Dutch context kindergartens are formally part of the primary school system rather than the ECEC system. Hence, issues related to school choice, funding, (structural) quality features etcetera, apply equally to kindergartens as to the other primary school grades.

The Dutch system can be described as a highly decentralized system, balanced by accountability (OECD 2016b; Ministry of Education, Culture and Science 2016). While both public and private schools exist (the former caters to around one third of primary school-aged children), almost all primary schools are publicly funded.[7] Private schools are generally based a particular religion (e.g. Protestant, Islamic) or educational model (e.g. Montessori). Every citizen has the constitutional right to start a private school and receive government subsidies. In fact, equal public funding rules apply for private and public schools.

Schools have a large degree of autonomy on how they allocate funding: a so-called 'lumpsum funding system' is in place. In addition to the number of pupils, funding rules take into account the composition of staff (funding increases with the age of teachers), the educational level of parents (schools with a high concentration of disadvantaged children receive more funding) and the school neighbourhood (schools in specific deprived areas, '*impulsgebieden'*, receive additional funding). Schools have a high degree of autonomy in how they allocate the public funding. For instance, the additional funding for a high share of disadvantaged children is not earmarked. Furthermore, schools also have a large degree of autonomy over the teaching content and method. There is, for example, no national curriculum: "'Freedom to organise teaching systems' means that both public and private schools are free to determine what is taught at schools and how this is taught, within legal boundaries." (Ministry of Education, Culture and Science 2016: p. 1)

The large degree of autonomy is balanced by national quality standards and examinations and a strong inspectorate of education. These standards include teacher qualifications, subjects to be taught, attainment levels but also the content of national examinations. The inspectorate of education plays an important role in monitoring schools. The inspectorate evaluates school quality and classifies schools performing below standard as (very) weak schools. Weak schools will be monitored closely; when they do not adhere to the standards in the following assessment funding will decline or schools will be closed.

3.2.3.2 Kindergarten

When children turn age four, they are entitled to enrol in kindergarten and almost all children do so. Participation is mandatory from age five. Given that they are part of the primary school system, kindergartens are free and universally accessible. Programs are for 25 hours per week, five days a week during school weeks. Kindergarten teachers are essentially primary school teacher in terms of qualification requirements (Bachelor degree) and wages. To facilitate parental employment children can use subsidized out-of-school care before or after the operating hours of kindergarten programs. There is no standard duration of kindergarten as children generally enrol during the school year (i.e. when they turn four). Formal schooling starts in 'group' 3 (Grade 1 in according to international definitions) and the transition to this grade depends on the teacher's assessment of the child's school readiness. This implies that some children may spend less than 18 months in kindergarten, whereas others may spend over three years in kindergarten. Grade retention in the second year of kindergarten ('*kleuterbouwverlenging*') occurs when the child is not considered 'school ready' and is relatively high (around 7%).

As mentioned above, targeted policies include early school programs for disadvantaged children. The total budget for these programs offered in kindergarten is smaller than the preschool budget (50 versus 200 million in 2010). However, in addition to these programs, the funding of

[7] Less than .5% of students are in private, not publicly funded primary schools (OCW 2016)

primary schools depends on the school's student composition. As kindergartens are part of primary schools, schools with a higher share of disadvantaged children receive more funding and can hire for instance more (specialized) kindergarten staff.

3.2.3.3 End of primary school

When children leave primary school around age 12, children enrol in a one of the three main secondary school tracks (see Section 3.2.4). Track choice is determined by the school track recommendation and an independent, standardized test. Scores of these tests correspond to a specific track recommendation. A recent reform has increased the relevance of the school recommendation. Before 2014–2015, teachers formulated their recommendation when the results of the standardized tests were available and in the large majority of cases the recommendation was consistent with the test results. Both the school recommendation and test results played a role in the actual track placement. Since 2014–2015, the school recommendation has become dominant for track placement. Teachers formulate their recommendation before test results are available; this recommendation should be reconsidered if test results indicate a higher track.

3.2.3.4 Inequality issues

The Netherlands has a universal kindergartens integrated in the primary school system. Evidence indicates beneficial effects of Dutch kindergartens, especially for disadvantaged children. Exploiting exogenous variation in kindergarten enrolment opportunities, Leuven et al. (2010) report that increasing the enrolment opportunities significantly improves achievement scores. Consistent with the literature on ECEC effects before kindergarten (van Huizen & Plantenga 2018), disadvantaged children benefit from increased kindergarten participation while non-disadvantaged children do not benefit. This finding implies that Dutch kindergartens substantially reduce achievement gaps: increasing the kindergarten enrolment opportunities by one month reduces the gap between disadvantaged and non-disadvantaged children by almost 10% (given a gap of .6–.7 SD).

A second important feature of the Dutch system is the rather high school segregation, which can probably be attributed to the high degree of parental choice and the high degree of school autonomy. In this policy context, higher educated parents appear to navigate carefully through the educational landscape. A comparison with the US indicates that segregation in the bigger cities in the Netherlands is at least as high as in large US cities (Ladd et al. 2011). Evidence also shows that segregation is increasing over time (Inspectorate of Education 2018). High SES families tend to opt for schools with specific educational programs (e.g. Montessori), that often ask for high voluntary parental fees.[8] Moreover, the existence of specific religious schools (e.g. Islamic schools) may contribute to ethnic segregation (Inspectorate of Education 2018). Nevertheless, descriptive evidence does not suggest clear negative consequences of segregation in terms of achievement (inequality). For instance, migrant achievement gaps declined between 1994 and 2004; SES gaps remained fairly stable (Ladd & Fiske 2011). Moreover, according to the most recent wave of PIRLS, the Grade 4 achievement gaps between

[8] Although Dutch schools are free, schools may ask for voluntary parental contributions to finance extra-curricular activities (e.g. school trips, music lessons). When families do not pay these contributions, their children can be excluded from extra-curricular activities but not from core education activities.

low and high achievers (10th versus 90th percentile) are the smallest in the Netherlands (UNICEF 2018).

3.2.4 Secondary school

3.2.4.1 A tracked system

The Dutch secondary education system contains three different tracks: 1) vmbo: pre-vocational secondary education. This track takes four years to complete and gives access to MBO. The vmbo track is divided into four different levels or subtracks; 2) havo: senior general secondary education. This track takes five years to complete and gives access to HBO (higher vocational training, leading to a Bachelor's degree); 3) vwo: pre-university education. This track takes six years to complete and gives access to WO (universities, leading to a Master's degree).

Some schools offer a combination of tracks (sometimes at different locations), other schools offer only one type ('categorial schools'), for instance only vmbo or vwo. The former type of schools generally offers more comprehensive classes in the 1st grade or first two grades of secondary schools, the so-called 'bridge classes'. These bridge classes consist of a combination of tracks. For instance, a common system is to offer a one-year vmbo/havo class and a two year havo/vwo class. The lowest achieving students in the first year vmbo/havo class will be tracked in a vmbo track, while the better performing students will move to a havo class or join the second year havo/vwo bridge class. Given that bridge classes are very common in the Netherlands, mobility in the first and second year is substantial: for most students the final track is clear in year three of secondary school. Around 25% moved to another track (Inspectorate of Education 2015).

3.2.4.2 Inequality issues

In general, low SES children and children with a migration background are more likely to enrol in a lower secondary school track. However, this can be expected as track placement in the Netherlands is to a large extent achievement based and these groups on average obtain lower achievement scores. However, gaps remain generally significant when conditioning on end of primary school standardized test scores (e.g. Inspectorate of Education 2018). Theoretically, the residual gap could be explained by measurement error in achievement, differences in non-cognitive skills, parental choice or discrimination. Hence, it is not obvious that this is evidence of inequality of opportunities. Studies relying on international data (e.g. PISA) produce mixed results on the effects of tracking (Brunello & Checchi 2007).

An interesting feature of the Dutch tracking system is the existence of bridge classes, as they facilitate mobility between tracks in the first year(s) of secondary school. Quasi-experimental evidence shows that a lack of opportunities to enrol in a bridge class (due to lower local supply of comprehensive schools) negatively affects the probability to complete higher education. This effect is larger for higher SES pupils (van Elk et al. 2011). Given this result it seems worrying that the use of bridge classes has declined in the past decade (Education Council 2018).

3.3 Data and methodology

3.3.1 (Pre-)COOL: Design and sample

The findings presented in this Chapter are based on the Dutch Pre-COOL and COOL data. Both datasets include a range of competence tests as well as information on family background. As in Chapter 2 (Germany) and 5 (UK), the analysis takes a long-term perspective (age 2–14) in the analysis of the extent and evolution of achievement gaps (RQ1). The analysis focuses on the early years (age 2–6) using Pre-COOL (2010/11–2014/15) and relies on data from the most recent wave of COOL (2013/14) to estimate the evolution of skill gaps until age 14. Hence, this approach combines a longitudinal cohort design (age 2–6, Pre-COOL data) with a pseudo-cohort design (age 5–14, COOL data). Our results for RQ1 are based on the full unbalanced sample.

The analysis explaining end-of-primary school achievement gaps by preschools achievement differences (RQ2) requires longitudinal data. These analyses therefore rely on the balanced sample. COOL is suited for this purpose as it follows children from kindergarten to the end of primary school.[9]

3.3.1.1 COOL$^{5–18}$

COOL$^{5–18}$, Cohort Study Educational Careers of 5/18-year-olds (*CohortOnderzoek OnderwijsLoopbanen*; COOL) has been collected triennially from school year 2007–2008 to 2013–2014.[10] The sampling design of COOL is school-based, implying that children are sampled from grades rather than cohorts (similar to NEPS, see Chapter 2). This means that the results are a better reflection of stage-specific inequalities than age-specific inequalities.

The data collection was executed by a larger consortium: ITS and the Kohnstamm Instituut were responsible for the part concerning kindergarten/primary schools, whereas Cito and GION were responsible for the part concerning secondary schools and MBO (age 14–15/17–18). COOL is based on a multi-cohort sequence design and includes in total 6 cohorts: 4 cohorts in COOL1, plus one additional kindergarten cohort in COOL2 and one in COOL3 (see Table 1).

The kindergarten and primary school data consist of a representative school sample (400 schools/almost 28,000 children in COOL1) and an additional sample of schools with a high share of children from disadvantaged backgrounds (150 schools in COOL1). Given the general aim of this report I rely on the representative sample: this sample cover around 6% of schools in the Netherlands. The collection of the representative sample took into account the national distribution of schools (using national administrative records) in terms of social-ethnic composition (6 categories), denomination (4 categories: public, Protestant, Catholic, other), province (12), degree of urbanization (5 categories). With respect to these four dimensions, the actual distribution of schools in the representative COOL sample does not significantly differ from the national distribution of schools.

Although the set-up of COOL is longitudinal, which allows children to be followed for several years, a substantial part of the sample drops out of COOL between consecutive waves. Around 44% of COOL2 Grade 3 children participated in COOL1 (kindergarten) and almost 60% of COOL3 Grade 6 children participated in COOL2 (Grade 3). There are three major reasons for panel attrition: schools do not participate in COOL anymore[11]; grade retention; or children move to another school.

Finally, in addition to kindergarten and primary school data, COOL contains information

[9] This is not (yet) possible with Pre-COOL data.

[10] See Driessen et al. (2009) for more information on the sampling procedures of COOL.

[11] For instance, around 40% of schools that participated in COOL1 did not participate in COOL2.

on achievement of children in the third year of secondary schooling (age 14–15). However, a sample-population comparison of the distributions in terms of school track, region and degree of urbanisation indicates that the secondary school data is not fully representative of the population (Zijsling et al. 2017). For instance, children in the vocational (vmbo) track are underrepresented (and those in the havo and vwo track overrepresented).[12] As no weights are provided it is not straightforward to correct this. The results on achievement gaps in secondary schooling should therefore be interpreted cautiously. COOL also includes achievement data for individuals aged 17–18. However, the timing of data collection beyond age 14 is track-specific and therefore no overall comparisons can be made for the oldest COOL group.[13] In this report I therefore rely on the age 5–14 data.

Table 1 Overview of the COOL waves, cohorts and samples.

			COOL1 07/08	COOL2 10/11	COOL3 13/14
Sample: number of schools					
Kindergarten/ primary schools			400	406	340
Secondary schools			81	151	107
Sample: number of children					
Stage	Age				
Kindergarten (year 2)	5	Cohort	C1-K07/08	C2-K10/11	C3-K13/14
		N	10069	9261	7279
Grade 3	8	Cohort	C1- K04/05	C2- K07/08	C3- K10/11
		N	9288	10109	7449
Grade 6	11	Cohort	C1-K010/2	C2- K4/5	C3-K07/08
		N	8545	9444 (12538)	7907
Secondary school year 3	14	Cohort	C1-K98/99	C2-K1/2	C3-K4/5
		N	8884	21384	16297

Notes: For RQ1 we rely mainly on COOL3. The analysis of RQ2 is based on the kindergarten 2007/08 cohort (cells shaded grey). Sample size refers to the representative sample; numbers indicate gross sample size.

[12] In COOL3, 38,6% of the sample follows a vmbo track, whereas this is almost 54% in the population.

[13] Age 18 for those in the vwo track (6th and final year of vwo), age 17 for those in the havo track (5th and final year of havo), and age 18 for those in mbo.

3.3.1.2 Pre-COOL

Pre-COOL is a cohort study that includes rich information on child development in the early years. The Pre-COOL sample includes in total over 3000 children who were age 2 in 2010 and consists of two subsamples: a center-based cohort and a home-based cohort.[14] Child development is assessed annually from age 2; average age was around 2.6 years at the first assessment in 2010–2011. For the analysis I use all five currently available data waves (2010–2011/2014–2015). Given our research design it is important to note that the fourth Pre-COOL wave (2013–14) overlaps with COOL3.

The existing COOL data collection infrastructure was taken into account when recruiting the Pre-COOL participants.[15] COOL schools that were willing to participate in Pre-COOL ('Pre-COOL schools') played a central role in the recruitment of both subsamples. First, the recruitment of the centre-based cohort occurred mainly via Pre-COOL schools. Schools were asked which ECEC providers (day care centres, playgroups) most children participated in before enrolling in the specific primary school. Pre-COOL centres are therefore generally located near Pre-COOL schools. Second, the home-based sample is based on a random sample draw by Statistics Netherlands in the postcode areas of Pre-COOL schools. The home-based cohort includes both children participating in ECEC and children not participating in ECEC. Additional recruitment effort was undertaken to increase the participation of children with migrant parents, including parents with a Moroccan and Turkish background.[16] Although Pre-COOL subjects are generally well-spread geographically (both rural and urban areas; all provinces are covered), Pre-COOL is not designed as a representative sample. However, as demonstrated below, after assigning weights based on the representative COOL sample, results from Pre-COOL are highly consistent with results from COOL.

3.3.2 Achievement measures

The central aim of this study is to examine how achievement gaps evolve over an extended observation window. For our main analysis I therefore rely on language and math achievement test data, as these are to a large extent comparable in the Pre-COOL (age 2–6) and COOL (age 5–14).

3.3.2.1 Pre-COOL

Pre-COOL contains two types of assessment data. First, a test battery has been developed for Pre-COOL to measure child development in the early years. These tasks were administered by trained research assistants. The rich data on the early years is one of the strong features of the Dutch data. Several domains have been assessed consistently from age 2 to 5 using the same task, increasing the difficulty with the age of the child.[17]

Second, for children aged 4 to 6 (kindergarten until 1st grade), Pre-COOL includes data

[14] In addition to the age 2 cohort, Pre-COOL includes an age 4 cohort. However, the latter sample is much smaller (around 700-800) and contains relatively limited child development measures.

[15] See Mulder et al. (2014) and Slot et al. (2015) for further details on the sampling procedures.

[16] For instance, families were visited at home by Berber and Arabic-speaking research assistants.

[17] For instance, receptive vocabulary has been tested every year from age 2 to 5 using a short version of PPVT.

from the Cito child achievement monitoring system (Cito-Leerlingvolgsysteem; Cito-LVS), administered by primary schools. In this monitoring system children can be tested twice a year (mid- and end-year test results are available). The Cito Language for Toddlers (Taal voor Kleuters) and Math for Toddlers (Rekenen voor Kleuters), the two tests administered in kindergarten, are both rather comprehensive tests. Cito Language for Toddlers measures language development (conceptual awareness, including receptive vocabulary) and emergent literacy (metalinguistic awareness).[18] Cito Math for Toddlers captures number knowledge, measurement and geometry. Both tests consist of two parts that are administered on two separate days and each part requires 20 to 30 minutes to complete. Several studies indicate good test reliability.[19]

Since the Cito monitoring system is used often by Dutch primary schools this data could often be supplied with limited additional effort from schools. However, not all Pre-COOL schools used this monitoring system and some schools administered the test only once or twice in kindergarten (there are four test moments in kindergarten). Test data is most complete for kindergarten mid-year 2; this is relevant as COOL also contains the mid-year Cito test results (see below).

For the analysis I use both Pre-COOL test battery data and Cito test data (see Table 2). For the description of the evolution of gaps over an extended observation window (RQ1) I focus on tasks that have been consistently administered from age 2–5 in the Pre-COOL test battery and that are related to competence assessments administered in COOL. For language, this concerns four tasks: PPVT (receptive vocabulary), phoneme task; nonword repetition task, grammar task. However, due to a rather low percentage of task completion of the latter two tasks, our main results are based on the first two tasks. Additional analyses include the nonword repetition task and grammar task. For math, I use the Cito Math for Toddlers tests (age 3–5).

3.3.2.2 COOL$^{5–18}$

Almost all tests used in COOL are from the Cito monitoring system.[20] Given that most schools use these tests and the associated software to process the test data, providing the test data to COOL implied marginal additional effort from these schools. If schools do not use the Cito monitoring system, the researchers provide the test. Given that using these tests is an integral part of most schools' policies and therefore common practice, non-response is low: depending on the wave and stage, 94–98% of children participate in at least one test. Non-response in test scores can be due to absence on the test day (due to illness) or because the child moved to another school between the date schools supplied enrolment information and the test date.

Table 2 provides an overview of the COOL test data and shows where COOL and Pre-COOL overlap. In the kindergarten and primary school phase, test data is collected for three stages: second year of kindergarten (K–2; age 5), Grade 3 (age 8) and Grade 6 (age 11). For the analysis of gaps in the language domain I use the following language test data:

---

[18] In year one of kindergarten the emphasis is on conceptual awareness.

[19] See Lansink and Hemker (2012) for more details on the Language for Toddlers test and Janssen et al. (2005) and Koerhuis and Keuning (2011) for more information on the Math for Toddlers test.

[20] An exception is the NSCCT (Non-School Cognitive Capacity Test; Niet-Schoolse Cognitieve Capaciteiten Test), a test that is similar to IQ test. This data is only available for children in Grade 3 and 9 (year 3 of primary school).

- Kindergarten (second year): Language for Toddlers (as in Pre-COOL, see above).
- Grade 3 and Grade 6: Vocabulary; Reading comprehension. In the main models I rely these three tests to generate a composite language measure.
- Grade 6: Cito End test, language component. This data is only available when schools administer this test as a regular part of school activities. In the relevant time period, around 85% of children made the Cito End test. Because some schools use alternative tests and the decision to do so is probably non-random, this sample may be not be fully representative (although our weighting strategy may to some extent account for this). I therefore consider the composite measure discussed above as the main measure for Grade 6 language achievement.
- Secondary school (third year): COOL includes a Math test (some use Math test developed by Cito); different versions with overlapping items depending on track: IRT equivalent scale.

For the description of gaps in the math/numeracy domain I use Cito Math for Toddlers (age 5; as in Pre-COOL, see above) and Cito Math tests for Grade 3 and Grade 6. Similar to the analysis of the gaps in the language domain, I present results using the math component of the Cito End test.

Table 2 Assessment data: extent and evolution of language and math gaps (RQ1).

Domain	Age	Measure/test	Data	
			Pre-COOL	COOL3
Language	2-6	Composite measure (Pre-COOL test battery): vocabulary, phoneme awareness, grammar *, verbal short-term memory (nonword repetition) *	X	
	4	Cito test: Language for Toddlers	X	
	5	Cito test: Language for Toddlers	X	X
	6	Composite measure (Cito): vocabulary, reading speed/accuracy	X	
	8/11	Composite measure (Cito): vocabulary, reading comprehension		X
	11	Cito End test – language component		X
	14	Composite measure (Cito): reading comprehension, grammar/orthography		X
Math	3	Cito Math for Toddlers (short version)	X	
	4	Cito Math for Toddlers	X	
	5	Cito Math for Toddlers	X	X
	8/11	Cito Math		X
	11	Cito End test – math component		X
	14	Math		X

Notes: * Given the rather low task completion rate, we exclude these from the main to remain. However, including these additional tests does not substantially change the results.

As most of the tests are part of the child achievement monitoring system, it is important to note that the test results play a crucial role in the educational careers of children. The school recommendation and end of primary school tests (in particular the Cito End test) determine whether someone has access to specific secondary school tracks. When formulating the school recommendation, teachers take into account these test results.

3.3.3 Socio-economic status and migration background

Socio-economic status (SES) of the child's family is measured using information on parental education.[21] This information is gathered via two sources in COOL: a survey directly administered to parents and school registry data. The survey information provides more detailed information on parental education. However, as the survey response rate is around 70%, relying only on the parental survey data implies a substantial drop in the number of observations. Moreover, the analytical sample may not be representative given that non-response may be non-random. A strong feature of COOL is that in addition to data from the parental survey, schools supply information on parental background. Schools often have this information as funding depends on the educational level of children enrolled in school (see Section 3.2.3). In fact, schools have a financial incentive to register this information, especially when the parental background is low. School registry data is less detailed but more complete (over 95% of the sample). Based on the parental background information supplied by schools it appears that survey non-response is indeed non-random: non-response is relatively low among low educated parents and migrant parents. Using both sources of parental education, the SES level for over 98% of the children in the COOL sample can be determined.

As in COOL, Pre-COOL also used multiple sources to complete data on family background: parent questionnaires, centre records and school registry data (as in COOL). The response rate for the first parent questionnaire was rather low (83% for the home-based cohort and 45% for the centre-based cohort).[22] Questions on the educational level of the parent were therefore also included in questionnaires of subsequent waves. When combining the information from different sources, a SES level can be assigned to about 80% of the children in the sample.

Following Bradbury et al. (2015) and the other chapters of this report, SES is measured by the highest qualification obtained by the child's parents. I use the information from the parental survey and complement this with data supplied by schools. In case information for one of the parents is missing, I use data on the highest level of education of the other parent. As in the other chapters of this report, I distinguish between three SES categories (see Table 3): 1) High: the required years of education is 15 or higher. At least one parent holds a bachelor's degree or a higher degree; 2) Medium: the required years of education is between 11 and 14. The lion's share of this group completed vocational education (mbo or equivalent); 3) Low: parents who completed the vocational track of secondary school (or lower educational attainment), requiring maximum 10 years of education. The distribution is roughly 43% (high), 40% (medium) and 17% (low).

Data on migration background was also collected through multiple sources. In addition to

---

[21] In both Pre-COOL and COOL indicators for household income are incomplete and imprecise.

[22] This difference can be explained by the fact that families participating in the home-based cohort were visited at home. In contrast, the questionnaires were distributed to the families by the centre and families were asked to return the completed questionnaire.

parent questionnaires and centre/school records, Statistics Netherlands provided information on the country of origin of families in the home-based cohort of Pre-COOL. To analyse the migrant-native gap in achievement I distinguish between children with no migration background (both parents were born in the Netherlands) and children with a migration background (at least one parent was born outside the Netherlands). Moreover, in additional analyses I also distinguish between children from families with a Turkish and a Moroccan background.

Table 3 Classification of SES groups by parental education.

SES Category	Answer category survey	School data	Required years of education
Low (≤10)	No	Max. lo	Max (mavo): 6+4=10
	lbo or similar	Max. lbo	
	mavo		
Medium (11-14)	havo/vwo or similar	Max. mbo	6+5==11 (havo); 6+6==12 (vwo)
	mbo or similar		6+4+2=12 (low level mbo); 6+4+4=14 (high level mbo)
High (≥15)	hbo	hbo/wo	6+5+4=15 (hbo); 6+6+4=16 (wo)

3.3.4 Methods

3.3.4.1 Standardisation and age corrections of achievement scores

In line with the general analytical strategy of this report I use a relative approach to measure achievement gaps. In some of the (Pre-)COOL analysis on language gaps I use composite measures, which are calculated by averaging z-scores of the relevant tests and subsequently standardising these z-score averages (as in Chapter 2 (Germany) and 5 (UK)).

In both Pre-COOL and COOL data there is substantial within-wave variation in the age of children. The aim is to remove this variation through residualisation (as in Bradbury et al. (2015); see also Chapter 2 and 6).[23] However, a problem arises when applying this approach to stages beyond kindergarten as grade retention is rather common in the Netherlands. This implies that in the primary and secondary school sample age is not only associated with test scores due to a pure age effect but also through a grade retention. In fact, whereas age is consistently positively associated with achievement until the end of kindergarten, test scores in Grade 3 and 6 generally appear to be negatively related to the child's age. These results suggest that age-variations in primary and secondary school grades capture mostly grade retention effects. Hence, residualising on age would in practice imply the exact opposite as controlling for a pure age effect.

3.3.4.2 The evolution of achievement gaps (RQ1)

To examine the extent and evolution of achievement gaps (RQ1) I estimate the relative position

[23] The main results are based on residuals from a regression of test scores predicted as a linear function of child age. Including higher order polinomials does not substantially change the results.

of the different groups using OLS, relying on the unbalanced sample and using in total 25 different dependent variables (15 tests for language and 10 for math, see Table 3). All models estimating SES effects control for migration background. The migrant-native gaps are estimated as total gaps as well as gaps net of SES. Because the trajectories are modelled for an extended observation window – from age 2 to 14 – important qualitative changes occur and (relative) gaps in achievement may not always be directly comparable over time. For example, in the primary school years reading comprehension is a relevant component of language skills, but this skill cannot be measured in the early years. The discussion of the results will therefore focus on different segments of the trajectories, mainly the early years (age 2–5) and the years from kindergarten to the end of primary school (age 5–11).

The analysis uses both cross-section weights and longitudinal (attrition) weights. Given that the unbalanced samples in COOL are representative, weights play a minor role in the models using COOL. No longitudinal weights are required for the unbalanced samples of COOL: there is panel attrition and refreshment from wave to wave and the unbalanced samples are representative. However, cross-section weights are calculated but these do not play a significant role for the results. The cross-section weights for COOL correct for the minor changes in the marginal distribution of the central variables (SES, migrant background) over time and between cohorts. For the main results, I consider the COOL3 kindergarten year 2 (age 5) – this is essentially the Pre-COOL cohort – as the base cohort and calculate weights for the older cohorts (age 8, 11, 14). More importantly, I use COOL3 distributions to calculate cross-section weights for the Pre-COOL sample. In this way, I correct for the overrepresentation of high SES in Pre-COOL. Next, to correct for panel attrition in the Pre-COOL sample, I apply inverse probability weighting. Wave 2 (age 3) of Pre-COOL is considered as the base wave: this wave has the largest number of observations as a significant number of additional children enrolled in the sampled ECEC centres after age 2. Interestingly, since age 5 data from COOL and Pre-COOL overlap, I am able to evaluate the extent of the potential remaining bias due to non-representativeness of the Pre-COOL sample. As discussed below, the Pre-COOL results are by and large consistent with the results from the representative sample, suggesting that the bias is negligible after applying the weighting strategy.

In addition to using cross-section and longitudinal weights, it is necessary to adjust standard errors for clustering given the sampling design of Pre-COOL and COOL (Abadie et al. 2017). Standard errors in the COOL analyses are clustered at school level. The standard errors in the models based on Pre-COOL are clustered at the target (Pre-)COOL school.

Finally, it should be noted that an important limitation of the approach is that I have to rely on data from older cohorts for the analysis beyond age 6: the scores for age 8 (Grade 3), age 11 (Grade 6) and Grade 14 (3rd year of secondary school) are for older birth cohorts.[24] Theoretically, age effects may be confounded with cohort effects. The results should thus be interpreted as predictions of the evolution of achievement gaps of the youngest cohort under the assumption that cohort effects do not play a major role. In Section 3.4.1 I discuss results indicating that estimates are comparable when the analysis relies on genuine cohort data (i.e. the K07/08 cohort, see Table 1).

[24] There is a cohort gap of around 9 years between the youngest cohort (Pre-COOL / COOL3-K13/14) and the oldest cohort (COOL3-K4/5), i.e. the youngest cohort was enrolled in kindergarten in 2013–2014 while the oldest cohort was enrolled in kindergarten in 2004–2005.

3.3.4.3 Explaining school gaps by preschool differences (RQ2)

To estimate the proportion of school gaps that can be attributed to achievement differences before enrolling in 1st grade, longitudinal data is required that follows children when they from preschool years through primary school. For the analysis of RQ1 I therefore use the COOL K07/08 cohort: these children were age 5 in the second year of kindergarten in the school year 2007–2008 and reached the end of primary school (6th grade) by the age of 11. The focus is on language skills (see Table 4) as these are more consistently assessed than math skills. The scores on the Language for Toddlers test, made in kindergarten, play a central role in the analysis as this test measures the initial achievement before moving to Grade 1 of primary school. Unfortunately, for the COOL K07/08 cohort no data on the Math for Toddlers is available. However, the children made an alternative test (an ordering/sorting), which I use in the empirical model (as discussed below). As I rely on the balanced sample, the number of observations is low (over 900 children) compared to the number of observations in the unbalanced. The balanced sample may not be representative if panel attrition is non-random. I correct for panel attrition using inverse probability weighting strategy. As in the models describing the evolution of gaps (RQ1), we cluster standard errors at the school level.

While the models relevant for RQ1 are estimated by OLS, an alternative estimation model is required for the analysis of RQ2. Since the analysis conditions on preschool achievement, OLS will produce biased estimates if the preschool achievement measure contains measurement error. There are several reasons why the test scores do not perfectly reflect the child's true language skills. First, the item composition of the test matters: children with the same skill level may obtain different test results as they are by chance more or less familiar with specific test items. Second, children often make guesses in tests and therefore have a non-zero probability of answering an item correctly. Third, the test day may affect scores: some students may for some random reason experience a 'bad day' and score below their potential on the test, i.e. they would have obtained a higher score if the test would have been administered on a different day. Finally, some students may exert more effort in the test than others and as a result attain a relatively high test result.

Measurement error in preschool achievement will lead to spurious regression to the mean. The intuition is that, on the one hand, children scoring in the top of the distribution are on average more likely to obtain a lower score in a subsequent measurement (i.e. some high achievers were lucky at the first test). On the other hand, children scoring in the bottom of the distribution are more likely to obtain a higher score in a subsequent measurement (i.e. some low achievers were unlucky at the first test). Moreover, it is likely that measurement error is systematically related to SES and migration background. This implies that measurement error in preschool achievement leads to an underestimation of the persistence of achievement and consequently to an underestimation of the share of school gaps that is attributable to preschool differences (and an overestimation of the additional SES effect in the school years). In line with other chapters in this report, we address this issue by applying an instrumental variable approach (see also Bradbury et al. (2015) and Jerrim & Vignoles (2013) for applications of this approach). The central idea is that in the first stage language achievement is predicted by an alternative achievement test made in the preschool years. The alternative tests used in the first stage is a kindergarten ordering/sorting test. The scores from this test are highly correlated with the kindergarten language tests. The main assumption is that the measurement error is uncorrelated between the achievement measures. This seems plausible as there is no overlap in the items and the tests are made on different days. Addressing measurement error appears to have important implications for the empirical results: in some specifications, the share of school gaps explained by preschool differences more than doubles after correcting for measurement error.

Table 4 Test data used for RQ2.

	Kindergarten (year 2) (age 5)	Grade 3 (age 8)	Grade 6 (age 11)
Test	Language for Toddlers	Language (Vocabulary; Reading comprehension)	Language (Vocabulary; Reading comprehension)

Notes: All tests used are developed by Cito.

3.4 Empirical results

3.4.1 The evolution of achievement gaps (RQ1)

3.4.1.1 Inequalities by socio-economic status

Figure 2 presents the evolution of language trajectories from age 2 to 14 by family SES, visualizing the results from 15 regression models, based on 23 tests and more than 40,000 observations. Overall, the pattern does indicate persistence or even diverging paths of different SES groups. Here I focus on three segments of the figure: the early years (age 2–5, based on the Pre-COOL test battery); kindergarten and Grade 1 (based on Pre-COOL Cito results); and kindergarten to secondary school (based on COOL3).

In the early years SES gaps in language achievement are already substantial. The high-low SES gap is around .65 SD at age 2 and increases in the following years: between age 2 and 3 the gap increases significantly to around .74. Between age 3 and 5 the high-low SES gap continues to widen at a pace of around .1 SD per year to .94. However, these increases are not statistically significant, implying that we cannot rule out stability of the SES gap in language achievement in the early years.[25] The finding that SES gaps are persistent or increase between age 2 and 5 appears not to depend on the specific language test: although gaps in vocabulary achievement are generally somewhat larger than gaps according to other language tests, the pattern over time is similar for all four Pre-COOL language tests (see Appendix 3.3).

Interestingly, the pattern of increasing gaps in vocabulary is not found when using raw (non-standardized) test scores or IRT-based test scores: in fact, these suggest convergence rather than divergence scores (see Leseman et al. 2017; Verhagen et al. 2016; see Appendix 3.1 for more details). However, as the aforementioned studies report no statistically significant change over time when using a similar indicator of SES (mother's level of education), one can argue that these results are overall consistent with the results presented in this report. Moreover, the difference can be explained by the fact that the variance of the relevant scores drops substantially from the first (age 2) until the fourth (age 5) Pre-COOL wave. As the focus of this

[25] This could be due to panel attrition of more than two-thirds of the sample from age 3 to 4 when children enrol in kindergarten. Given the relatively small sample size, the power of test is rather low.

report is on the development of relative achievement gaps, I adjust for changes in the variance of the test score distribution. These results highlight that absolute and relative gaps should not be confused.

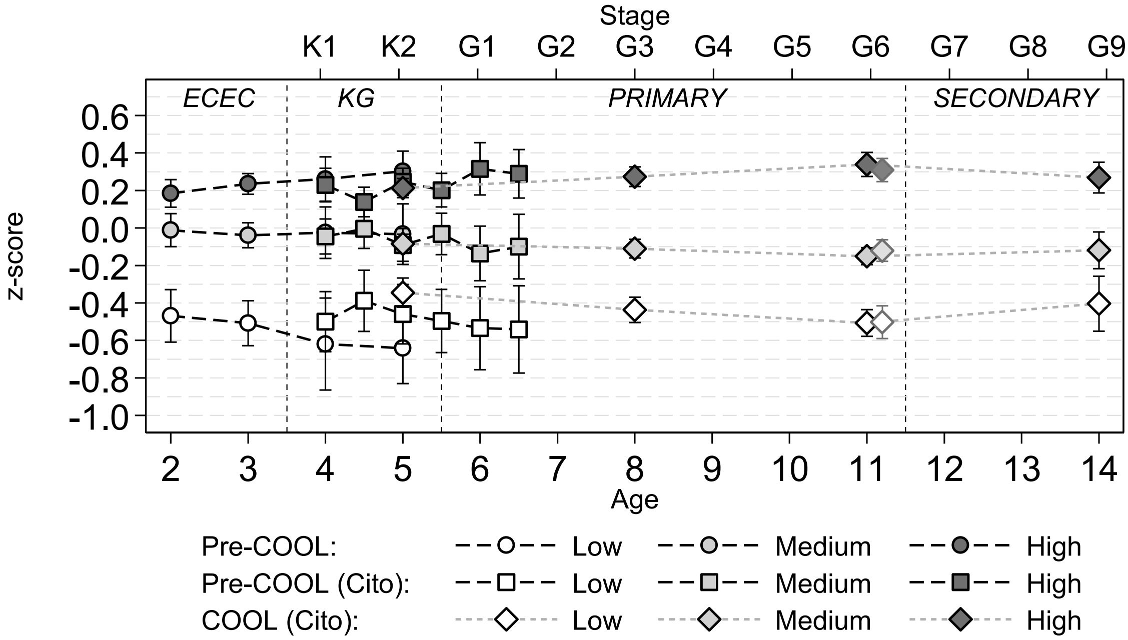

Figure 2 Language z-scores of children by SES (parental education).

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (language) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

The findings concerning the Pre-COOL Cito language test indicate persistence of the gap in the first years after kindergarten and school enrolment. The analysis based on the largest sample (age 5; mid-year 2 of kindergarten) indicates a high-low SES gap of around .7 and a medium-low SES gap of about half that size. Compared to the Pre-COOL test battery results, the gaps based on this test are somewhat smaller. Nevertheless, the estimates for high and medium SES children are almost identical when considering the same time periods. For the low SES group, there is a visible but small difference and the confidence intervals overlap. Interestingly, the Pre-COOL Cito and and COOL Cito results, both based on the Language for Toddlers test, are almost identical. However, the Pre-COOL low SES point estimate is slightly lower than the COOL counterpart. This suggests that the Pre-COOL results are highly comparable to those based on the representative COOL sample. However, low SES may be more negatively selected into the Pre-COOL sample, implying that models based on Pre-COOL probably somewhat overestimate SES language gaps in the population.

The third relevant segment of Figure 2 describes the evolution of SES gaps in language achievement when children move from kindergarten through primary school. Again, the estimates point out divergence of trajectories, although high-low SES gaps grow at a slower pace in this phase than in early childhood: from .56 SD in kindergarten to over .85 by the end of primary school (i.e. around half a SD per year). Yet, low and medium SES children follow parallel pathways, indicating that the main development over time can be characterized as high SES children moving away from the rest of the population. Interestingly, the results based on the language component of the Cito End test are highly consistent with our main end of primary school estimates (the former indicates a .81 low-high SES gap).

Next, Figure 2 does indicate a pattern of convergence after leaving primary school. However, one may argue that this result can be explained by the use of a different composite language measure (vocabulary and reading comprehension versus grammar/orthography and reading comprehension). Nevertheless, even when I use tests measuring the same domain (i.e. test data on reading comprehension is available for both stages), I find a reduction of the SES gap after enrolling in secondary school (from around .8 to .7 SD). As mentioned earlier, it should be stressed that the secondary school sample is not fully representative, and this result should therefore be interpreted cautiously.

Finally, it should be noted that the results beyond age 7 are all based on COOL3. This dataset concerns older cohorts – e.g. children in the Grade 6 sample were born around 6 years earlier than the children in the sample used to describe the development of gaps in the early years. In additional analyses I have used data representing a single cohort (using all three waves of COOL) to estimate the age 5-11 trajectories. The results can be found in Appendix 3.2. The estimates clearly show that there are hardly any cohort differences in stage-specific positions and trajectories. Although the future development of the trajectories of the Pre-COOL cohort remains of course uncertain to some extent, these results indicate that cohort effects are probably negligible when the birthyears from the different cohorts are not too far apart.

While Figure 2 provides evidence on the evolution of language trajectories, Figure 3 represents the results of an equivalent analysis for math achievement scores. The figure represents the results from 10 regressions using in total over 37.000 observations on the development of math trajectories by SES from age 3 to 14. As in language domain, math gaps are already substantial in the early years before entry into kindergarten. The patterns for math indicate a larger degree of stability: between age 3 and 5, the high-low gap hovers around .7; and the medium-low gap stays close to half of that gap. Although the age 3 gap in math is similar to the age 3 gap in language achievement, the math gap is more persistent in the early years. Moreover, between kindergarten and the end of primary school SES trajectories diverge to some extent: the high-low SES gap increases from .56 to .73 SD. Compared to the results for language, this change is rather small (i.e. less than .03 SD per year).[26] The gap remains very stable as children leave primary school and move through the first years of secondary school. Again, there appears to be hardly any difference between the age 5 Pre-COOL and COOL results, suggesting that the Pre-COOL results are representative.[27] Finally, additional results also indicate negligible cohort differences in math trajectories, which once more indicates that the presented results beyond age 6 are plausible predictions of the development of the Pre-COOL cohort.

[26] Results from pooled regressions show that this increase in the SES math gap is not statistically significant.

[27] As mentioned above, the age 5 test data are not perfectly comparable between COOL1 and COOL3 as a different test was used.

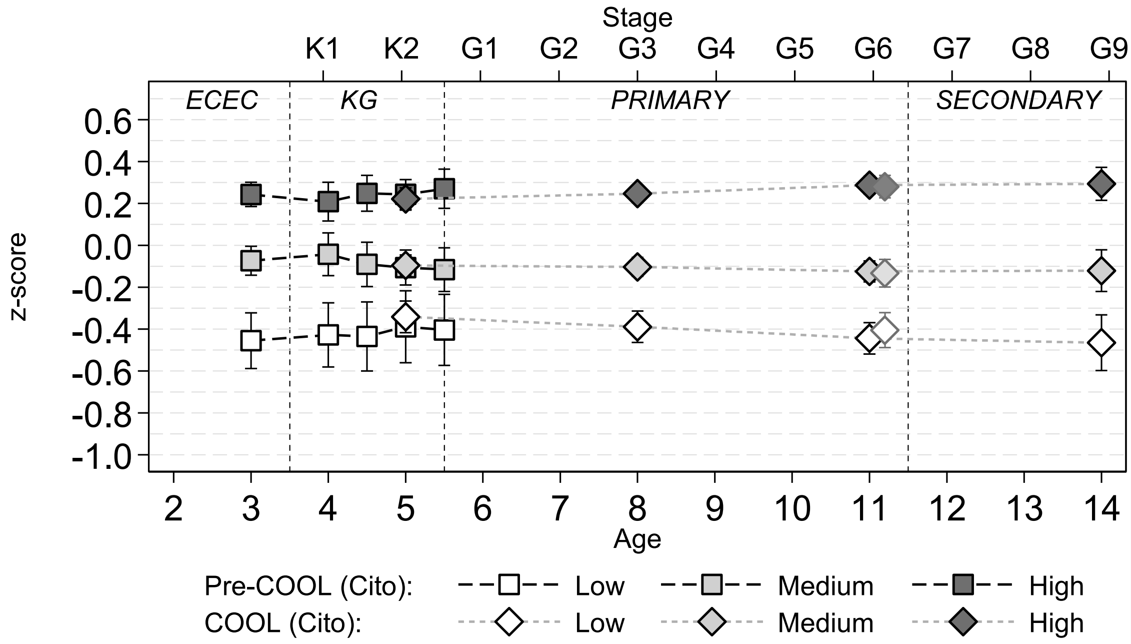

Figure 3 Math z-scores of children by SES (parental education).

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

3.4.1.2 Inequalities by migration background

Figure 4 plots the results from models estimating migrant-native gaps in language achievement from age 2 to 14. Overall, the figure indicates that gaps close, especially in early childhood. As in the discussion on SES trajectories, I discuss three separate segments of the figure. First of all, the results from Pre-COOL show that language achievement differences between children with and children without a migration background are sizeable before children enrol in kindergarten (.85–.9 SD). SES differences can only account for a small share of this gap: after controlling for SES, the migrant-native gap at age 2–3 remains substantial (around .77 SD). However, more than half of this gap disappears after children move into kindergarten. Next, estimates based on the Pre-COOL Cito language test show that migrant-native gaps in kindergarten are substantial in kindergarten (.48–.59 SD for mid-year 2 kindergarten, depending on whether SES differences are accounted for).[28] However, the figure also shows that gaps continue to narrow during the period in kindergarten.

Furthermore, the total (net of SES) gaps decline during the primary school phase by 18% (38). Estimates based on the Cito End test data point out an even more dramatic narrowing of the gap: the total (net of SES) language gap reduces by about 50% (75). No significant migrant-native gaps are found for children in year 3 of secondary school, suggesting a further convergence of trajectories after the transition to secondary school. Hence, the estimation results from the three separate segments of Figure 4 consistently point out that the migrant-native gap declines

[28] The kindergarten results from Pre-COOL Cito are again highly consistent with the COOL results.

between age 2 and 14. While the migrant-native gap is large before school entry, it appears that migrants catch up in primary and secondary school.

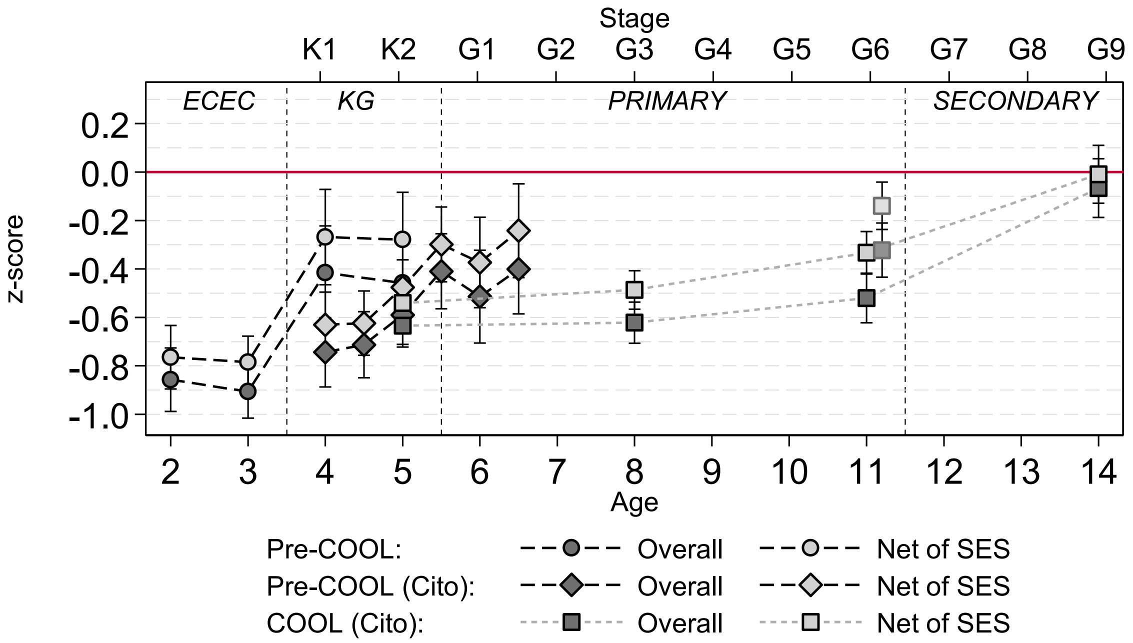

Figure 4 Total and net (SES) migrant-native gaps in language achievement.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

Although migrants appear to catch up in general, Figure 5 (total gap) and 6 (gap net of SES) indicate that the extent and evolution of language achievement gaps depend on the country of origin of the child's parents. The results clearly show that children with a Turkish background consistently and considerably lag behind.[29] Children with a Turkish background consistently score lower than children with a Moroccan background and other migrant groups. The total Turkish-native gap in language is especially large in the early years (1.5–1.9 SD at age 2–3), but the gap appears to narrow in the early years. However, the results based on COOL indicate that the Turkish-native gap in language achievement is also large in kindergarten: almost 1 SD (.8 SD net of SES). Moreover, children with a Turkish background show no signs of catching-up in language as the total gap hovers around 1 SD in Grade 3 and Grade 6.[30] Although less pronounced, language achievement differences between children from Moroccan families and children with native Dutch parents are also large in the early years (over 1 SD at age 2–3). However, for this group the migrant-native gap narrows somewhat during the school years; this catching-up pattern is stronger when conditioning on SES.

---

[29] Note that the Pre-COOL results are based on a relatively small number of observations; migrant group-specific results should therefore be interpreted cautiously. Results from COOL (5-11) are based on much larger samples and are therefore more reliable.

[30] Given that the COOL secondary school sample is not fully representative and children with a Moroccan and Turkish background are underrepresented, no secondary school gaps are presented.

Overall, from age 2 until the end of primary school, children with a Moroccan and Turkish background are in a disadvantaged position. Although these groups catch up to some extent in the preschool and kindergarten stage, the language achievement penalties appear to be rather persistent once children enter primary school. The results concerning the Turkish-native gap are striking as they point out large gaps in the preschool period and almost no closing of the gap in the primary school phase.

Do migrant-native gaps in math achievement evolve according to similar patterns? Figure 7 shows the evolution of migrant-native gaps in math achievement over the educational career. While the migrant-native gap in math is sizeable and significant in the early years (the age 3 total gap is .56 SD; the net of SES gap .43 SD), the gap in math is consistently smaller than the migrant-native gap in language (Figure 4). Moreover, the evidence points out that migrants catch up in math during primary school: by the end of primary school the migrant-native gap net of SES is small (less than .1 SD) and only marginally significant ($p < .05$). Results from the math component of the Grade 6 Cito End test in fact indicate no significant difference in math achievement scores. Interestingly, the results show an (insignificant) migrant premium for this test when conditioning on SES. This evidence is important from a life course perspective, as Cito End test scores are relevant for the track placement in secondary school.

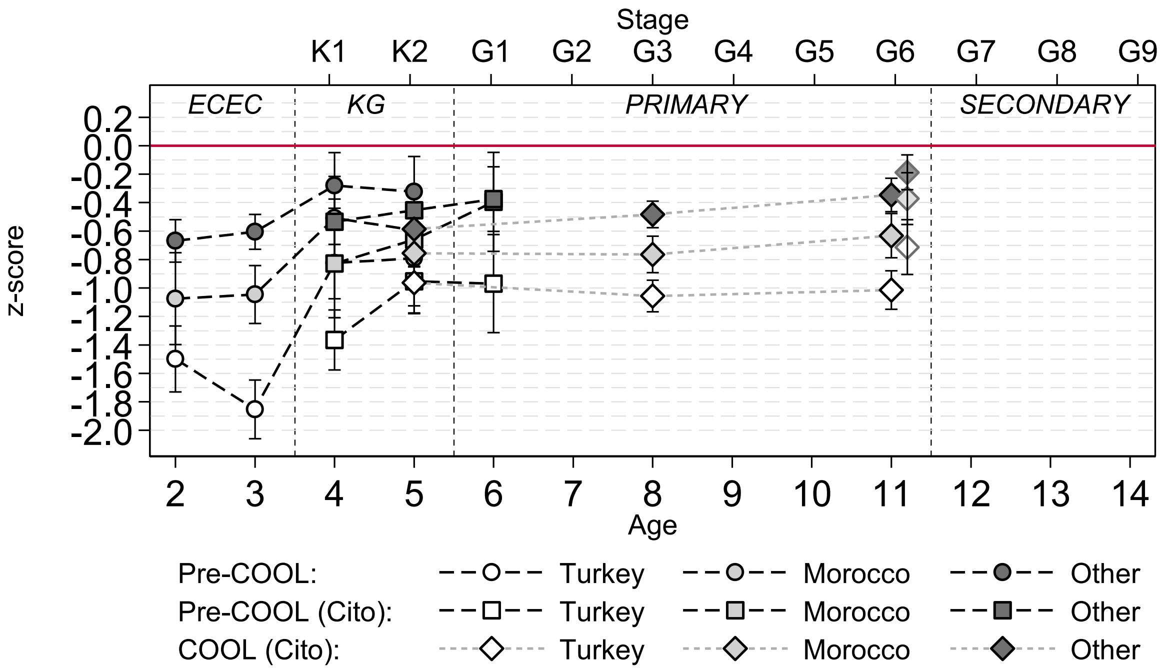

Figure 5 Total migrant-native gaps in language achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

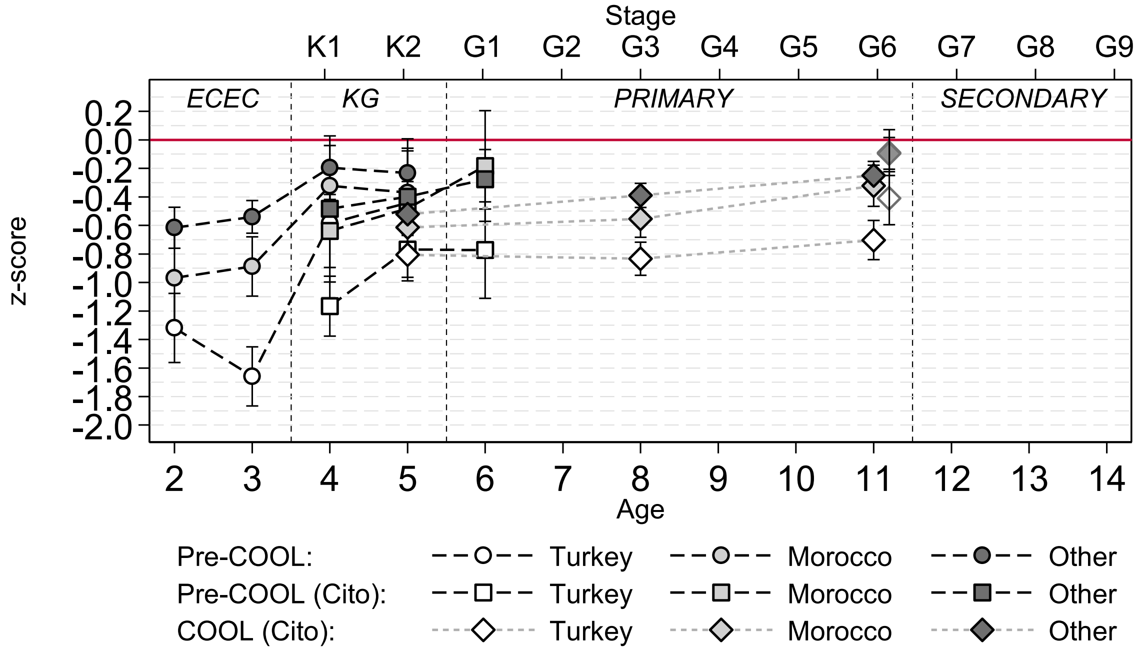

Figure 6 Net (of SES) migrant-native gaps in language achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 7 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

Figure 8 and 9 present the estimation results of math gaps by the country of origin of the child's parents. As in the language domain, children with a Turkish background perform worse in math than natives and other migrant groups in the early years: at age 3, the total Turkish-native gap in math achievement is around 1 SD. Nevertheless, the evolution of math gaps differs from the evolution of language gaps. In general, Turkish and Moroccan children catch up significantly during the primary school period (e.g. the Turkish-native gap decreases by around two thirds). It is striking that these two groups catch up more rapidly than other migrant groups. While in the preschool period other migrant groups on average outperform Turkish and Moroccan children, the results from models conditioning on SES point out that by the end of primary school the gaps are actually smaller for Turkish and Moroccan children. Consequently, no significant Turkish-native and Moroccan-native achievement gaps in math are found in Grade 6 when controlling for SES differences. The results based on the math component of the End Cito test show that children with a Turkish background score significantly lower than natives, although this gap is relatively small (.18 SD). However, for this test the Moroccan-native gap is not significant. Moreover, when conditioning on SES, the sign of the gap reverses, indicating an insignificant Turkish and a significant (.2 SD) Moroccan math premium.

All in all, the migrant-native gap in math scores are substantial in the early years but narrow considerably during primary school. The evolution of math trajectories of Turkish and Moroccan children is striking as they catch-up more than other migrants and the gap observed by the end of primary school is rather small and can be fully explained by SES.

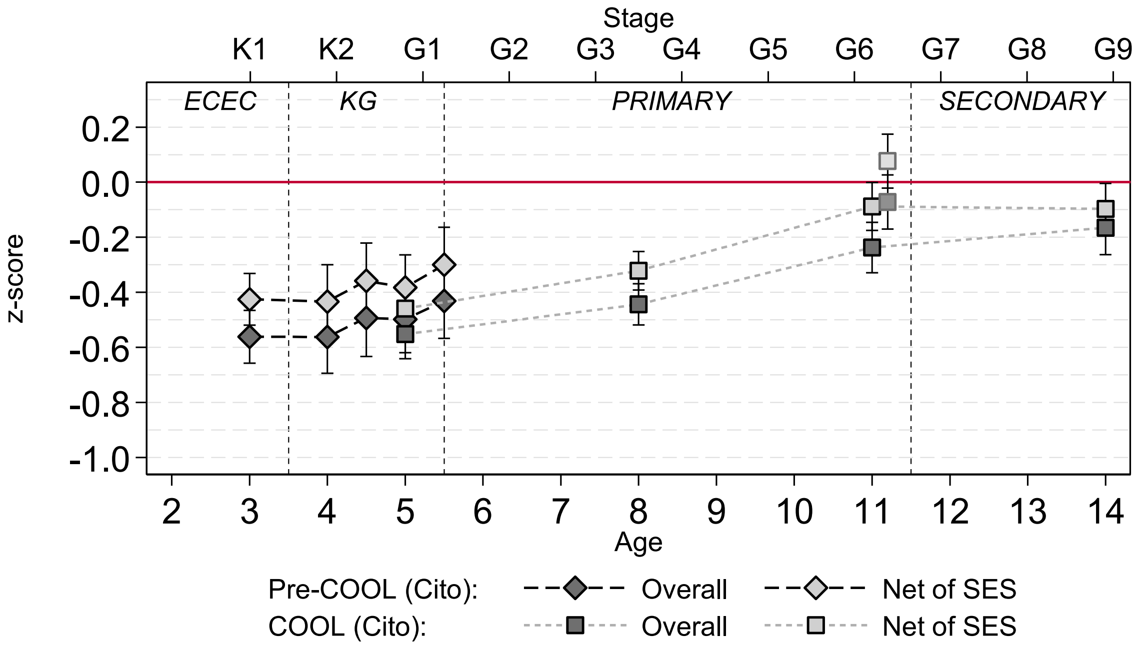

Figure 7 Total and net (SES) migrant-native gaps in math achievement.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

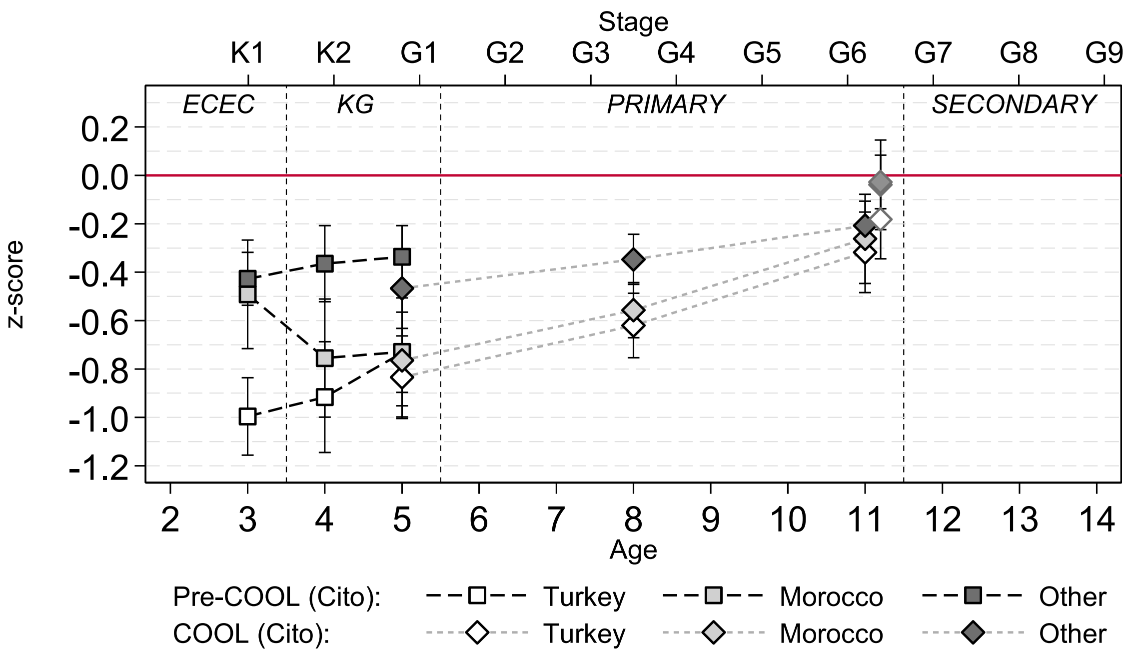

Figure 8 Total migrant-native gaps in math achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

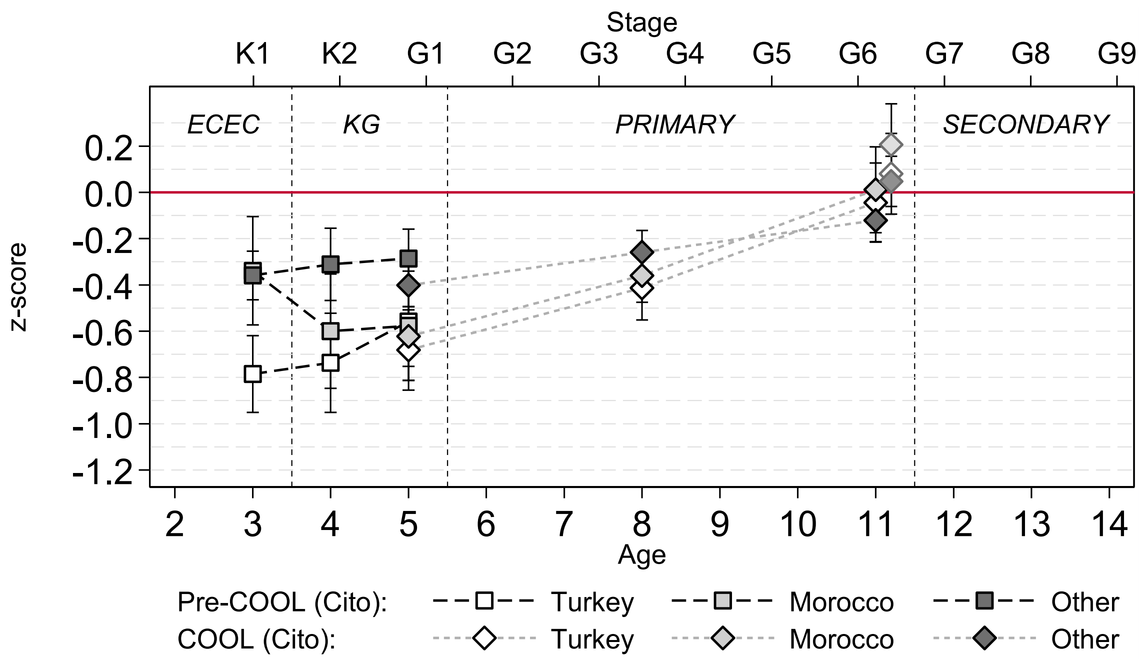

Figure 9 Net (of SES) migrant-native gaps in math achievement by country of origin.

Notes: Predictions are from stage-specific regression models. Grey-bordered diamonds represent estimates based on Cito End test (math) scores (age 11). Long-dashed black lines connect data within the same cohort: results before age 6 are based on the same cohort; results beyond age 7 are based on older cohorts. Capped spikes indicate 95% confidence intervals. Stage: K = Kindergarten, G = Grade level in school.

3.4.2 Explaining school gaps by preschool differences (RQ2)

3.4.2.1 Inequalities by socio-economic status

The proportion of school gaps attributable to preschool differences

The results presented in Section 3.4.1.1 provide evidence that SES gaps widen as children move through school. To examine the extent to which social inequalities in educational achievement in school can be explained by preschool inequalities (RQ2) I compare aggregate SES gaps with gaps according to common trajectories. Figure 10 presents average (observed) and (predicted) common trajectories (CT) of children from low, medium and high SES families. The average trajectories are based on a balanced sample, using COOL1–3. In general, these results are similar to the results presented in Figure 2 (unbalanced sample, using COOL3) and in Appendix 3.2 (unbalanced sample, using COOL1–3). As described in Section 3.4.1.1 for the unbalanced sample, in the balanced sample SES gaps also increase during the primary school years. This appears to be mainly the result of high SES children moving away from the rest of the population: the high-low gap increases from around .5 in kindergarten to around .8 in the final grade of primary school.

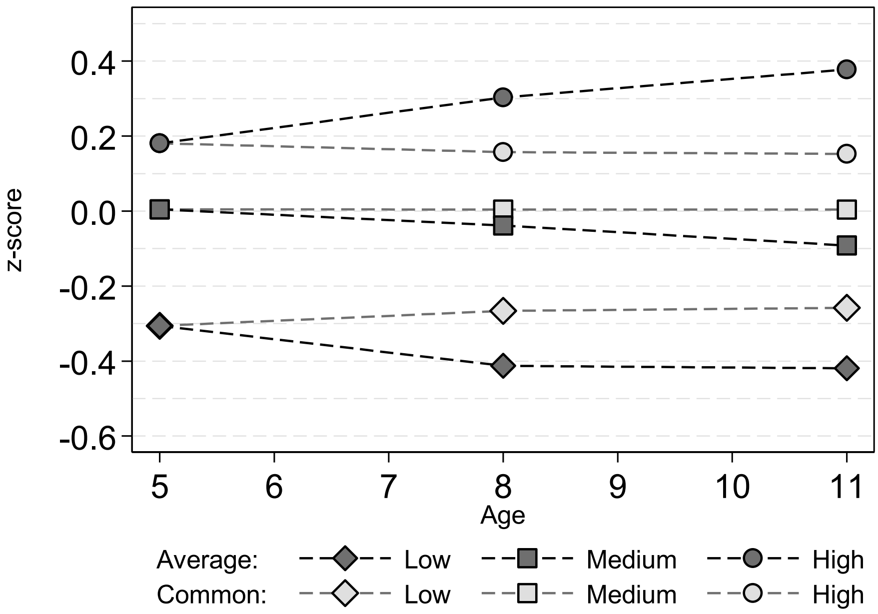

Figure 10 Aggregate and common trajectories by SES.

The common trajectories shows the predicted path of different SES groups under the assumption that all three SES groups would have developed in the same way after kindergarten. As discussed above, these IV estimates are correcting for measurement error. The common trajectories represent a natural counterfactual condition of no SES effects in primary school. Given regression to the mean (i.e. due to transitory shocks), one can expect that the end of school gaps are smaller according to the common trajectory model than the actual preschool gaps. This implies that if SES had no additional effect in primary school, the aggregate gaps in school can be expected to be smaller than the preschool gaps.

Whereas the aggregate SES gaps in the final grade of primary school are .33 SD (medium-low) and almost .8 SD (high-low), these gaps are predicted to be .26 and .41 respectively in the absence of additional SES effects. This suggests that 80% of the medium-low gap and 52% of the high-low gap can be attributed to differences in preschool achievement. Hence, comparing the aggregate and common trajectories, one can conclude that the majority of language achievement gaps observed in the final grade of primary school can be explained by preschool inequalities.

Instead of comparing aggregate with common trajectories, I can estimate SES gaps in language achievement in primary school while conditioning on preschool achievement levels (again using an IV approach): see Table 5 for the estimation results. The findings show that achievement scores are highly persistent: a 1 SD increase in preschool scores is associated with a .8 SD increase in scores obtained in Grade 6. When comparing aggregate gaps with gaps conditioning on preschool achievement level, it appears that 75% of the medium-low and 48% of the high-low gap can be explained by differences in preschool achievement. Overall, the estimated percentage explained by preschool achievement is comparable (though slightly lower) with the estimates based on the aggregate-common trajectory comparisons discussed above.

Table 5 SES gaps in language achievement and the role of preschool differences (RQ2).

		Stage (approx. age)			
		Grade 3 (age 8)		Grade 6 (age 11)	
		OLS	IV	OLS	IV
SES (ref. low)					
	Medium	0.374**	0.116	0.327**	0.081
		(0.146)	(0.139)	(0.141)	(0.164)
	High	0.716***	0.311**	0.797***	0.411**
		(0.193)	(0.147)	(0.191)	(0.178)
Language score KG			0.832***		0.792***
			(0.097)		(0.107)
N		870	870	870	870
Low-Medium gap					
% preschool			*69*		*75*
% additional			*31*		*25*
Low-High gap					
% preschool			*57*		*48*
% additional			*43*		*52*

Notes: Clustered standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Diverging trajectories

While the evidence indicates that preschool achievement gaps are rather persistent, there appear to be significant additional SES effects beyond the preschool years. I examine whether these additional SES effects depend on the child's position in the kindergarten achievement distribution: are high SES children more likely to recover from a bad start in kindergarten than low SES children? Does the position of low SES children who are low achievers at age 5 worsen or improve during the primary school years? The predictions based on IV models allowing for the interaction between SES and age 5 achievement level are presented in Figure 11. The figure shows predicted trajectories of low, medium and high SES children who performed low (–1 SD above the average), average or high (+1 SD above the average) in a language test in kindergarten. It should be noted that these results should be interpreted as simulated rather than actual trajectories and that the point estimates are surrounded by large confidence intervals.

The figure clearly shows that achievement of high SES children systematically – i.e. independent of the initial position – improves more than the position of low and medium SES children. The predictions also suggest that low–achieving high SES recover to a large degree from their initial disadvantage between kindergarten and 6th grade (from –1 SD to –.6 SD). The position of medium SES remains quite stable during the school years and this holds for low, average and high achievers.

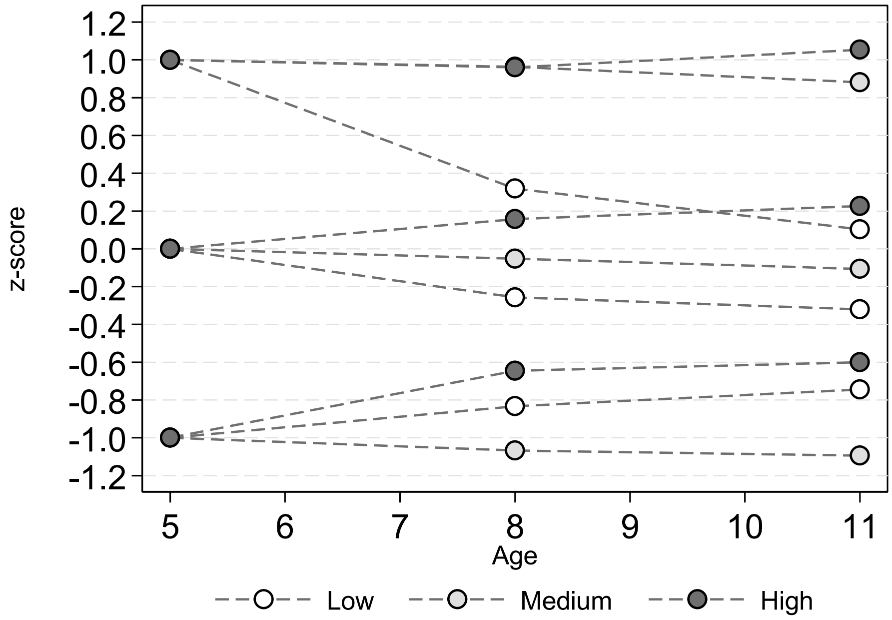

Figure 11 Diverging trajectories by SES.

The most remarkable results concern low SES children in the bottom and top of the kindergarten achievement distribution. The figure shows that low achieving, low SES children are not being left behind but catch-up somewhat (i.e. by more than .2 SD) during the primary school period. The point estimates actually indicate that this group outperforms low achieving, medium SES children. Nevertheless, regarding children in the top of the kindergarten achievement distribution, low SES children appear to lag behind their medium and high SES peers. Particularly among high achievers in kindergarten, the increase in the SES gap is substantial: by the end of primary school, the position of low SES children starting in the top of the achievement distribution is overtaken by high SES children with an average kindergarten achievement level. In general, these estimates suggest that achievement is less persistent among low SES children – i.e. they have a stronger tendency to move towards the mean of the achievement distribution. Although we cannot identify the factors that cause these patterns, two important features of the Dutch system may provide an explanation for the striking results. First, school segregation by socio-economic background is relatively high in the Netherlands, which may lead to downward mobility of high-achieving, low SES children. Second, schools with high concentrations of disadvantaged children receive additional funding to improve the overall quality, which may boost upward mobility of low-achieving, low SES children.

Implications for secondary school track placement

The maximum longitudinal component of COOL is 6 years, implying the data can be used to follow children from preschool to Grade 6 (the end of primary school), but not into secondary school. The role of preschool achievement in secondary school track enrolment can therefore not directly be tested. However, the data does contain information on variables that determine track placement in secondary school. First, track placement is based on the results of a standardized test. In the relevant time period, around 85% children participated in the Cito End test. The standardized test scores correspond to a specific track recommendation (the "test track recommendation"). Second, the school recommendation plays an important role. School recommendations take into account achievement scores but also other characteristics of the child such as motivation and attitudes towards school.

Appendix 3.4 replicates the main findings presented in Table 5 using Grade 6 Cito End test results. Preschool language scores (net of SES) appear to be strong predictors of these test results: a 1 SD increase in preschool achievement is associated with a .7 SD increase in the Cito End language score and a .9 SD increase in the Cito total score. As a result, the test track recommendation is also significantly predicted by preschool achievement: a 1 SD increase in preschool achievement increases the probability to receive a high test track recommendation by over 30% (see Appendix 3.4).

Consistent with many other studies (e.g. Inspectorate of Education, 2018), significant SES effects on Cito End test scores and track recommendations are found. The SES effect on the test track recommendation is slightly smaller than the SES effect on the school track recommendation. The results point out that a sizeable share (30–50%) of the SES effect on Cito End test scores is explained by differences in preschool scores. Consequently, preschool language achievement explains a substantial proportion of the SES effect on the probability to receive a high track recommendation (defined as a havo or vwo track recommendation): 29 (high-low gap) to 67% (medium-low gap). Similarly, the SES effect on the school track recommendation can, to a significant degree, be attributed to preschool differences in language achievement: 27 (high-low gap) to 51% (medium-low gap). Hence, the results imply that SES inequalities in the two central dimensions determining track placement can be partially explained by achievement differences in kindergarten.

3.4.2.2 Inequalities by migration background

The proportion of school gaps attributable to preschool differences

As discussed in Section 3.4.1.2, the evolution of aggregate migrant-native trajectories in language achievement shows some signs of convergence: gaps appear to narrow as children move through primary school. Despite the process of convergence, migrant-native gaps remain sizeable and significant even in the last grade of primary school. To estimate the share of migrant-native school gaps that can be attributed to preschool differences, we follow an equivalent approach as in 3.4.2.1. First, I compare aggregate and common trajectories (see Figure 12). The common trajectories represent the counterfactual trajectories of no additional migrant effects in the primary school period. According to this approach, the migrant-native gaps by the end of primary school can be entirely explained by preschool differences if actual gaps by the end of primary school would be identical to those predicted by the common trajectory models. The results show that the total migrant-native gap according to the common trajectories would decline from .85 in kindergarten to .68 in Grade 6 (Figure 12, left panel). In the absence of additional migrant penalties during school the achievement gap would narrow as a result of regression to the mean (note that the IV approach controls for "spurious" regression to the mean due to measurement error). Nevertheless, the reduction of the aggregate total migrant-native gap appears to be even larger than predicted by the common trajectories model. This pattern is also found when conditioning on SES differences (Figure 12, right panel). Hence, the aggregate-common trajectory comparison implies that differences in preschool achievement fully account for the migrant-native gap in school.

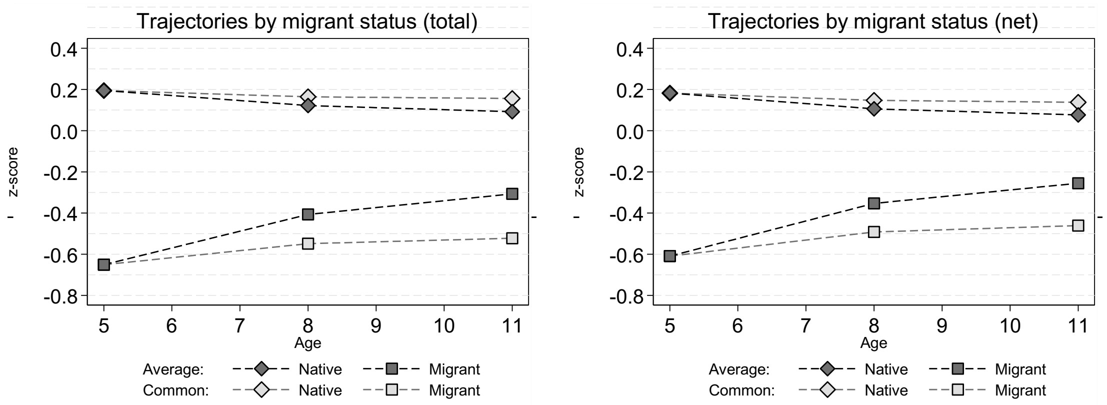

Figure 12 Aggregate and common trajectories by migration background.

Table 6 presents the results of IV models estimating migrant-native gaps in language achievement in primary school. Whereas aggregate gaps are significant in Grade 6 (total: .4 SD; net of SES: .33 SD), the coefficient indicating the migrant-native gap becomes positive (though generally not significant) when controlling for differences in preschool achievement. This means that migrants entering school with the same language achievement as their native counterparts would experience a migrant premium rather than a penalty. Again, this result points out that migrants catch up during the primary school phase and that migrant-native gaps in language achievement can be entirely explained by preschool differences.

Table 6 Migrant-native gaps in language achievement and the role of preschool differences.

	Stage (approx. age)			
	Grade 3 (age 8)		Grade 6 (age 11)	
	OLS	IV	OLS	IV
Panel A: Overall gaps				
Migrant (ref. native)	-0.529***	0.207	-0.399**	0.315
	(0.156)	(0.173)	(0.151)	(0.225)
Language score KG		0.870***		0.844***
		(0.102)		(0.111)
N	870	870	870	870
Panel B: Gaps net of SES				
Migrant (ref. native)	-0.459***	0.200	-0.332**	0.296
	(0.160)	(0.156)	(0.139)	(0.198)
Language score KG		0.832***		0.792***
		(0.097)		(0.107)
N	870	870	870	870

Notes: Clustered standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Implications for secondary school track placement

On average, migrants obtain lower primary school achievement scores. Results based on COOL also show that migrants perform worse on the Cito End test (Appendix 3.5). The migrant-native gap is rather small and not always statistically significant. Given that secondary school recommendation is to a large extent based on achievement scores, a migrant penalty on secondary school recommendation can be expected. The estimation results indeed show that this is the case (see Appendix 3.5): migrants receive a lower test track recommendation (because they perform worse on the Cito End test) as well as a lower school track recommendation. Compared to natives, migrants face a 23 (test) and 16% (school) lower probability to receive a high track recommendation. Conditioning on SES reduces this gap substantially though. Furthermore, the point estimates indicate that migrants are more likely than natives to receive a high track recommendation when controlling for preschool differences in language achievement. Interestingly, this migrant premium is stronger for the school than for the test track recommendation. A potential explanation is that school track recommendations are not only based on past and current achievement levels, but also take into account future improvements of children's achievement levels.

3.5 Conclusions

This chapter combines data from two related longitudinal studies – COOL and Pre-COOL – to examine the early roots of social and migration-related inequalities in the Netherlands. The findings reported in the chapter document how language and math gaps evolve from early childhood (age 2–3) to adolescence (age 14). During this extensive observation window, children make several important life transitions and experience important qualitative changes in development.

The evidence clearly shows that SES gaps in language and math are already large before children enrol in kindergarten. SES gaps in language appear to be larger than in math though. The analysis of the evolution of gaps over an extended observation window (age 2–14) show that SES gaps are persistent or widen somewhat during the school years. In general, divergence is more pronounced in the language domain. The evidence indicates that a substantial share of language achievement gaps in school can be attributed to differences in preschool achievement. Concerning gaps measured in the final grade of primary school (Grade 6), 75 to 80% of the medium-low SES gap can be explained by preschool differences in achievement. A smaller share of the high-low SES gap is attributable to preschool differences (around 50%), which can be explained by the result that high SES children appear to move away from the rest of the population. Overall, the findings presented in this chapter imply that at least half of the SES gaps in school can be explained by differences that were already present before school entry.

Interestingly, simulated trajectories indicate that, depending on their preschool achievement, low SES children experience either a rise or fall of their relative position as they move through primary school. On the one hand, low SES children starting in the bottom of the achievement distribution preschool appear to catch up more than their medium SES counterparts and nearly as much as their high SES counterparts. On the other hand, low SES children who are in the top of the achievement distribution before school entry almost completely lose their initial advantage by the time they reach the final grade of primary school. The high degree of both downward and upward among low SES children may be due to the Dutch 'high segregation, high compensation' context: while school segregation is relatively high in the Netherlands, schools with a higher concentration of disadvantaged children receive significant amounts of additional funding.

The chapter also provides evidence of significant migrant-native gaps, which are particularly pronounced in the early years. As for SES inequalities, migrant-native gaps are larger in language than in math achievement. The extended observation window (age 2–14) results show a striking pattern of a narrowing migrant-native gap. Migrant and native trajectories converge more in the math domain; language gaps are overall more persistent. In general, there is strong evidence pointing out that migrants catch up in the early years and during primary school. The findings indicate that the migrant-native gap observed in primary school can be fully attributed to preschool differences in achievement. Compared to natives with the same preschool language achievement, migrants obtain an (insignificant) achievement premium.

Focusing on different groups of migrants, the results show that Moroccan and Turkish migrants perform worse than natives and other migrants in both domains. During the entire observation window, Turkish migrants consistently appear to be the most disadvantaged group. Moroccan children catch up somewhat in the language domain, but Turkish children do not seem to recover from their initial disadvantaged position in language achievement during the school years. However, there is compelling evidence indicating that both groups catch-up almost completely in the math domain.

There is now mounting evidence that the early years are important for later school achievement, educational attainment, labour market success and other relevant life outcomes (e.g. Doyle et al. 2009; Heckman & Mosso 2014; Kautz et al. 2014). This chapter further quantifies the relevance of the early years by describing the evolution of gaps from early childhood to adolescence and by estimating the role of preschool differences for later school achievement levels. The empirical results presented here once more highlight that the early years matter. Targeted high-quality early childhood education and care programs have the potential to significantly reduce achievement disadvantages in the early years (Elango et al. 2015). Moreover, since universal (high-quality) early childhood education and care programs often generate significant benefits only for lower SES children (van Huizen & Plantenga 2018), universal programs have also the potential to level the playing field. Although increased investments in such programs will involve substantial public short-term costs, various studies indicate that the societal long-run benefits will probably outweigh these costs (Elango et al. 2015; van Huizen et al. 2018). The reported evidence provides a strong rationale to prioritize programs and policies that effectively reduce achievement disadvantages in the early years.